

Repetitive Transients Extraction Algorithm for Detecting Bearing Faults*

Wangpeng He¹, Yin Ding^{† 2}, Yanyang Zi^{3 4}, and Ivan W. Selesnick²

¹*School of Aerospace Science and Technology, Xidian University, Xi'an, China*

²*Tandon School of Engineering, New York University, 6 Metrotech Center, Brooklyn, NY, USA*

³*State Key Laboratory for Manufacturing and Systems Engineering, Xi'an Jiaotong University, Xi'an, China*

⁴*School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an, China*

August 9, 2016

Abstract

This paper addresses the problem of noise reduction with simultaneous components extraction in vibration signals for faults diagnosis of bearing. The observed vibration signal is modeled as a summation of two components contaminated by noise, and each component composes of repetitive transients. To extract the two components simultaneously, an approach by solving an optimization problem is proposed in this paper. The problem adopts convex sparsity-based regularization scheme for decomposition, and non-convex regularization is used to further promote the sparsity but preserving the global convexity. A synthetic example is presented to illustrate the performance of the proposed approach for repetitive feature extraction. The performance and effectiveness of the proposed method are further demonstrated by applying to compound faults and single fault diagnosis of a locomotive bearing. The results show the proposed approach can effectively extract the features of outer and inner race defects.

*Preprint

[†]email: yd372@nyu.edu

1 Introduction

Rolling bearings are one of the most prevalent components in rotating machines and reciprocating machines [35]. Vibration-based fault detection has become the preferred technique for bearing fault diagnosis [44, 18]. Bearing vibrations are usually modeled as cyclostationary signals (or group-sparse periodic) [9, 10, 4]. When a localized defect occurs on the bearing, repetitive transients will be generated due to the passing of rollers over the defect [26, 38]. These transients have repetitive structure and are usually submerged in background noise. Many denoising methods have been introduced to extract fault features for the purpose of detecting faults in machines, such as wavelet transform [43, 12], singular value decomposition (SVD) [27], time-frequency analysis methods [21, 23], empirical mode decomposition (EMD) [30], methods for blind extraction of a cyclostationary signal and spectral kurtosis (SK) [9, 10, 5]. If compound faults exist, then the observed vibration signals are rather complex and it is difficult to identify each fault using traditional signal processing methods.

A number of approaches have been developed for the multiple fault or compound fault diagnosis. Principal component analysis (PCA) based method has been used to classify faults [29, 31, 33]. Support vector machine (SVM) based methods have been used for multi-fault diagnosis [46, 1] as well. Some other techniques such as neural network and independent component analysis (ICA), have been introduced to assist fault detection and classification [42, 8, 7]. Some of these methods require large data collection as a training set and further require an off-line training phase; and some methods rely on careful selection of features (e.g., wavelet packet sub-bands).

Adopting sparsity in the field of fault detection was initially illustrated in Ref. [45], where basis pursuit denoising (BPD) [14] was used to exploit sparse features in various domains to detect faults. Some recent works consider other sparse representations for fault diagnosis [44, 16, 25, 17]. The fault frequencies of potential bearing fault features can be simply obtained using the geometry of the components in many cases or directly obtained from the user operation manual. Many works have considered the fault frequencies as priori information [44, 32, 40, 34, 24].

In this work, a method using the temporal periodicity (namely fault characteristic or fundamental frequency) directly in the time domain is proposed to extract fault features while simultaneously denoising. The proposed method is based on convex optimization using non-convex regularization. Specifically, this paper aims to address the problem of extracting compound features caused by faults in vibration signals, where the features exhibit repetitive group sparsity.

In particular, the observed signal is modeled as

$$y = x_1 + x_2 + w, \quad (1)$$

where w denotes additive white Gaussian noise (AWGN), and x_1 and x_2 are both repetitively group-sparse signals with periods T_1 and T_2 respectively. Note that here the repetitively group-sparse signal means the useful features (group-sparse) appear repetitively. In other words, the features have group-sparse property and appear repetitively. To avoid confusion, here the “period” is referred to as the cycle for cyclostationary signal in Ref. [2]. In the case of compound faults detection of bearings, useful features x_1 and x_2 also satisfy the following two conditions.

1. The periods T_1 and T_2 are different.
2. Each period is not close to an integer multiple of the other.

Note that, in real applications, the fault features are not strictly group-sparse periodic. In other words, the real fault frequencies have a variation from the calculated frequencies (up to 1-2%), i.e., there are not strict periods T_1 and T_2 [35]. In such cases, the proposed approach can still work owing to the overlapping of group structure, which will be shown in the following sections. The proposed approach is not an improvement of the cutting edge techniques developed in the cyclostationary framework, instead, it is an alternative to these techniques. A work closely related to the proposed approach is the periodic overlapping group sparsity (POGS) problem [24], which assumes only one repetitive group-sparse component is present in the vibration signal. Notice that although the observed signal is modeled with two components, the proposed method also works when there exist only one fault component. Therefore, the proposed method generalizes POGS, and is useful when there exists multiple components.

Another related work is group-sparse signal denoising (GSSD), which is also known as overlapping group sparsity (OGS) with non-convex regularization [13], where mathematical derivations and proofs have been given in detail to show that non-convex regularization can be used to promote group-sparsity, while maintaining convexity of the problem as a whole.

The method proposed in this paper also uses the concept of morphological component analysis (MCA) [39], which is a method to decompose signals based on sparse representations. In contrast to MCA, the proposed method does not utilize any transform (e.g., Fourier or wavelet transform), i.e., the sparse features are in the signal domain. Moreover, non-convex regularization is used to

strongly induce sparsity while maintaining convexity of the proposed problem formulation. The proposed approach reduces to the OGS method without prior knowledge, i.e., we can utilize the sparsity-based OGS approach if we do not have prior knowledge of the characteristic frequencies.

2 Preliminaries

2.1 Notation

In this paper, the elements of a vector x are denoted as x_n or $[x]_n$. The norms of x are defined as

$$\|x\|_1 := \sum_n |x_n|, \quad \|x\|_2 := \left(\sum_n |x_n|^2 \right)^{1/2}. \quad (2)$$

A function of x determined by parameter a is denoted as $f(x; a)$, and to distinguish it from a function with two ordered arguments, e.g., $f(x, y)$.

2.2 Review of majorization-minimization

In this paper, the majorization-minimization (MM) approach is used to derive a fast-converging algorithm. This subsection briefly describes the MM approach for minimizing a convex cost function. The MM is an approach to simplify a complicated optimization problem into a sequence of simpler ones [22]. More specifically, consider an optimization problem

$$u^{\text{opt}} = \arg \min_u F(u). \quad (3)$$

Using MM, the problem can be solved iteratively by

$$u^{(i+1)} = \arg \min_u F^{\text{M}}(u, u^{(i)}), \quad (4)$$

where $F^{\text{M}} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is an upper bound (majorizer) of the objective function F , satisfying

$$F^{\text{M}}(u, v) \geq F(u), \quad F^{\text{M}}(u, u) = F(u). \quad (5)$$

Note that the majorizer $F^{\text{M}}(u, v)$ touches $F(u)$ for $u = v$, as shown in (5). Figure 1 illustrates the majorizer (red line) of a penalty function (blue line), which will be described in the following subsection. The proof of convergence for MM has been given in Ref. [28, Chapter 10]. More details

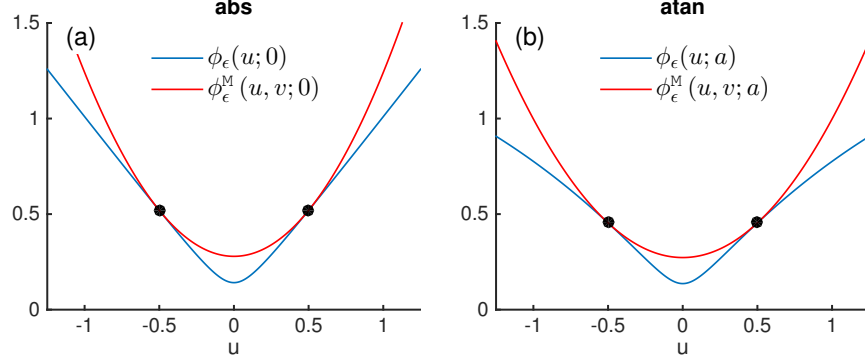


Figure 1: (a) Smoothed ℓ_1 -norm (abs) penalty function (blue) and majorizer (red). (b) Smoothed arctangent (atan) penalty function and majorizer (gray). For all the figures, the parameters are set to $v = 0.5, \epsilon = 0.02, a = 0.5$.

about the MM procedure can be found in [28, 22] and references therein.

2.3 Non-convex penalty functions

Non-convex penalty functions can promote sparsity more strongly than convex penalty functions [36, 37]. This subsection briefly describes the non-convex penalty functions which will be used in the proposed approach. The smoothed non-convex penalty function $\phi_\epsilon : \mathbb{R} \rightarrow \mathbb{R}_+$ is used in this work. Table 1 gives several examples of the functions defined by

$$\phi_\epsilon(u; a) := \phi(\sqrt{u^2 + \epsilon}; a), \quad \epsilon > 0, \quad (6)$$

where ϕ is a non-smooth penalty function satisfying the following properties:

1. $\phi(u; a)$ is continuous on \mathbb{R} .
2. $\phi(u; a)$ is twice continuously differentiable on $\mathbb{R} \setminus \{0\}$.
3. $\phi(u; a)$ is even symmetric: $\phi(u) = \phi(-u)$.
4. $\phi(u; a)$ is increasing and concave on \mathbb{R}_+ .
5. $\phi(u; a) = |x|$ when $a = 0$.

Note that, for both ϕ and ϕ_ϵ , the parameter $a \geq 0$ controls the concavity of the function.

The parameter ϵ controls the smoothness of the functions. As a special case, when $\epsilon = 0$, $\phi_\epsilon(u; 0) = \phi(u)$, then the penalty function is non-differentiable at 0. In practice, ϵ is specified very small, e.g. 10^{-10} , so that the function is differentiable. Figure 1 gives two specific examples of ϕ_ϵ .

A majorizer of penalty function ϕ_ϵ is given by

$$\phi_\epsilon^M(u, v; a) := \frac{u^2}{2\psi(v; a)} - \underbrace{\left(\frac{v^2}{2\psi(v; a)} - \phi_\epsilon(v; a) \right)}_{\text{only depends on } v}, \quad (7)$$

where $\psi(v; a)$ is listed in the third column of Table 1. Note that $\psi(v; a) > 0$ for all $v \in \mathbb{R}$. Also note that $\phi_\epsilon^M(u, v)$ is quadratic in u . In Ref. [19], a detailed proof has been given to show that when ϕ satisfies the above properties, the majorizer (7) satisfies the condition (5) for ϕ_ϵ .

2.4 Review of POGS

The periodic overlapping group sparsity (POGS) problem [24] considers the signal model

$$y = x + w, \quad (8)$$

where x is a repetitive group-sparse signal, and w is additive white Gaussian noise. The POGS method extracts x by solving a convex problem

$$x^{\text{opt}} = \arg \min_x \left\{ P_0(x) = \frac{1}{2} \|y - x\|_2^2 + \lambda \Phi(x, b; a) \right\}, \quad (9)$$

where $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}$ is defined as

$$\Phi(x, b; a) := \sum_n \phi_\epsilon \left(\left[\sum_k [b]_k [x]_{n+k}^2 \right]^{1/2}; a \right). \quad (10)$$

The function Φ is a regularization term that promotes repetitive group-sparsity. Moreover, $b \in \{0, 1\}^K$ is a binary weight array designated according to the period. To simplify following derivation, denote that

$$\Phi(x, b; 0) = \Phi(x; b), \quad (11)$$

where the penalty function is strictly convex as shown in Figure 1(a).

3 Repetitive transients extraction algorithm

In this section, the proposed algorithm termed repetitive transients extraction algorithm (RTEA) is presented.

Table 1: Sparsity-promoting penalty functions.

Penalty	$\phi(u; a)$	$\phi_\epsilon(u; a)$	$\psi(u; a) = u/\phi'_\epsilon(u; a)$
abs ($a = 0$)	$ u $	$\sqrt{u^2 + \epsilon}$	$\sqrt{u^2 + \epsilon}$
log	$\frac{1}{a} \log(1 + a u)$	$\frac{1}{a} \log(1 + a\sqrt{u^2 + \epsilon})$	$\sqrt{u^2 + \epsilon} \left(1 + a\sqrt{u^2 + \epsilon}\right)$
rat	$\frac{ u }{1 + a u /2}$	$\frac{\sqrt{u^2 + \epsilon}}{1 + a\sqrt{u^2 + \epsilon}/2}$	$\sqrt{u^2 + \epsilon} \left(1 + a\sqrt{u^2 + \epsilon}/2\right)^2$
atan	$\frac{2}{a\sqrt{3}} \left(\tan^{-1} \left(\frac{1 + 2a u }{\sqrt{3}} \right) - \frac{\pi}{6} \right)$	$\frac{2}{a\sqrt{3}} \left(\tan^{-1} \left(\frac{1 + 2a\sqrt{u^2 + \epsilon}}{\sqrt{3}} \right) - \frac{\pi}{6} \right)$	$\sqrt{u^2 + \epsilon} \left(1 + a\sqrt{u^2 + \epsilon} + a^2(u^2 + \epsilon)\right)$

3.1 Problem formulation

To extract two repetitive group-sparse components, an optimization problem is formulated as

$$\{x_1^{\text{opt}}, x_2^{\text{opt}}\} = \arg \min_{x_1, x_2} \left\{ P(x_1, x_2) = \frac{1}{2} \|y - (x_1 + x_2)\|_2^2 + \lambda_0 R(x_1, x_2; a_0) + \sum_{i \in \{1, 2\}} \lambda_i \Phi(x_i; b_i) \right\} \quad (12)$$

where $R : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is defined as

$$R(x_1, x_2; a_0) := \sum_n \phi_\epsilon \left(\left[\sum_{k=0}^{K_0-1} [x_1 + x_2]_{n+k}^2 \right]^{1/2}; a_0 \right). \quad (13)$$

The function R is an overlapping group sparsity (OGS) regularization function with group size K_0 ¹. There are two more regularizers in (12) promoting the repetitive group-sparsity of x_1 and x_2 respectively, and formulated in (11).

Furthermore, in problem (12), $b_1 \in \{0, 1\}^{K_1}$ and $b_2 \in \{0, 1\}^{K_2}$ are two binary-weighting arrays as

$$b_i = [\underbrace{1 \ 1 \ \cdots \ 1}_{N_{i1}} \ \underbrace{0 \ 0 \ \cdots \ 0}_{N_{i0}} \ \cdots \ \underbrace{1 \ 1 \ \cdots \ 1}_{N_{i1}} \ \underbrace{0 \ 0 \ \cdots \ 0}_{N_{i0}} \ \underbrace{1 \ 1 \ \cdots \ 1}_{N_{i1}}], \quad (14)$$

spanning M_i periods

for $i = 1, 2$ and M_1 and M_2 defines the number of periods included in b_1 and b_2 respectively.

Moreover, in contrast to MCA, which has two regularizers, problem (12) has three. The regularization term R is introduced because according to the signal model, the summation of the two components is also sparse. Note that, $(x_1 + x_2)$ might be sparse when x_1 and x_2 are not, but in

¹Using the notation of [24], it is denoted $\sum_i \phi_\epsilon(\|x\|_{i, K_0}; a_0)$.

(12), the regularizers with Φ does force x_1 and x_2 to be sparse.

3.2 Convexity of the objective function

The idea of using non-convex regularization with “maximizing concavity” in a convex problem has been illustrated in [36], wherein the quadratic formulation of data fidelity term can be used to compensate the non-convexity in the regularization so that the objective function is still convex. In this work, we adopt this idea for an extraction algorithm with simultaneous denoising. Entirely there are three regularizers in (12), and we allow one of them to be non-convex to promote the global sparsity more strongly. Moreover, the specific condition to assure the convexity of problem (12) is derived as the following proposition.

Proposition 1. *Suppose the parameterized penalty function ϕ_ϵ is defined by formula (6) and $\lambda_0 > 0$. If*

$$0 \leq a_0 < \frac{1}{K_0 \lambda_0}, \quad (15)$$

then the objective function $P : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ in (12) is strictly convex.

A proof of the above proposition is given in Appendix A.

4 Algorithm derivation

In this section, an algorithm is derived to solve (12) based on MM. The majorizer of R in (13) is $R^M : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$, and written explicitly as

$$\begin{aligned} R^M(x_1, x_2, z_1, z_2; a_0) \\ = \sum_n \left\{ r_0(n, z_1 + z_2) ([x_1]_n^2 + [x_2]_n^2 - [z_1 - z_2]_n [x_1]_n - [z_2 - z_1]_n [x_2]_n) \right\} + C(z_1, z_2), \end{aligned} \quad (16)$$

where $C(z_1, z_2)$ is a constant only dependent on z_1 and z_2 . In (16), r_0 is a function $r_0 : \mathbb{Z} \times \mathbb{R}^N \rightarrow \mathbb{R}$ given by

$$r_0(n, z_1 + z_2) = \sum_{j=0}^{K_0-1} \psi^{-1} \left(\left[\sum_{k=0}^{K_0-1} [z_1 + z_2]_{n-j+k}^2 \right]^{1/2}; a_0 \right). \quad (17)$$

The majorizer of function Φ has been derived in Ref. [24, Section 3.3]. Here, it can be rewritten

Table 2: Explicit steps of proposed method (RTEA).

Input: $y \in \mathbb{R}^N$, K_0 , λ_0 , λ_i , $b_i \in \{0, 1\}^{K_i}$, for $i \in \{1, 2\}$.
Initialization: $a_0 \geq 0$, $a_i \geq 0$, $x_i = y$, for $i = 1, 2$.
Repeat:
$[r_0]_n = \sum_{j=0}^{K_0-1} \frac{1}{\psi_0\left(\left[\sum_{k=0}^{K_0-1} [x_1 + x_2]_{n-j+k}^2\right]^{1/2}; a_0\right)}$
$[r_i]_n = \sum_{j=0}^{K_i-1} \frac{[b_i]_j}{\psi_i\left(\left[\sum_{k=0}^{K_i-1} [b_i]_k [x_i]_{n-j+k}^2\right]^{1/2}; a_i\right)}, \text{ for } i \in \{1, 2\}$
$[p_i]_n = 2 + 2\lambda_0[r_0]_n + \lambda_i[r_i]_n, \text{ for } i \in \{1, 2\}$
$[q_1]_n = y_n + (1 + [r_0]_n)[x_2 - x_1]_n$
$[q_2]_n = y_n + (1 + [r_0]_n)[x_1 - x_2]_n$
$[x_i]_n = [q_i]_n/[p_i]_n, \text{ for } i \in \{1, 2\}$
Until convergence
Return: x_1, x_2

using a slightly different notation, that $\Phi^M : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is

$$\Phi^M(x, z; b, a) = \frac{1}{2} \sum_n r(n, z) x_n^2 + C(z), \quad (18)$$

where $r : \mathbb{Z} \times \mathbb{R}^N \rightarrow \mathbb{R}$ is defined as

$$r(n, z) := \sum_{j=0}^{K-1} \frac{[b]_j}{\psi\left(\left[\sum_k [b]_k [z]_{n-j+k}^2\right]^{1/2}; a\right)} \quad (19)$$

Using the above results, the majorizer of the objective function P in (12) can be written as

$$\begin{aligned} P^M(x_1, x_2, z_1, z_2) &= \frac{1}{2} \|y - (x_1 + x_2)\|_2^2 + \frac{1}{2} \|(x_1 - z_1) - (x_2 - z_2)\|_2^2 \\ &\quad + \lambda_0 R^M(x_1, x_2, z_1, z_2; a_0) + \sum_{i \in \{1, 2\}} \lambda_i \Phi^M(x_i, z_i; b_i, a_i), \end{aligned} \quad (20)$$

where the function $R^M : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is a majorizer of R in (13), wherein the derivation in detail is given in Appendix B.

Note that in (20), an extra term $\frac{1}{2} \|(x_1 - z_1) - (x_2 - z_2)\|_2^2$ is introduced, which does not break

the property of majorizer, but helps to cancel the term $x_1^\top x_2$ in the data fidelity term. Then $P^M(x_1, x_2, z_1, z_2)$ can be written as

$$\begin{aligned} P^M(x_1, x_2, z_1, z_2) = & \sum_n \frac{1}{2} [p_1(z_1, z_2) \odot x_1]_n^2 + \frac{1}{2} [p_2(z_1, z_2) \odot x_2]_n^2 \\ & - [q_1(z_1, z_2) \odot x_1]_n - [q_2(z_1, z_2) \odot x_2]_n + C(z_1, z_2), \end{aligned} \quad (21)$$

where \odot denotes element-wise multiplication, and the coefficients $p_1, p_2, q_1, q_2 \in \mathbb{R}^N$ can be written explicitly as

$$[p_1(z_1, z_2)]_n = 2 + 2\lambda_0 r_0(n, z_1 + z_2) + \lambda_1 r_1(n, z_1) \quad (22a)$$

$$[p_2(z_1, z_2)]_n = 2 + 2\lambda_0 r_0(n, z_1 + z_2) + \lambda_2 r_2(n, z_2) \quad (22b)$$

$$[q_1(z_1, z_2)]_n = y_n + [1 + r_0(n, z_1 + z_2)][z_1 - z_2]_n \quad (22c)$$

$$[q_2(z_1, z_2)]_n = y_n + [1 + r_0(n, z_1 + z_2)][z_2 - z_1]_n. \quad (22d)$$

Note that r_1 and r_2 are functions defined by (19) dependent on the binary weighting arrays b_1 and b_2 in (14) respectively.

Consequently, using MM, the problem (12) is solved iteratively by

$$\{x_1^{(i+1)}, x_2^{(i+1)}\} = \arg \min_{x_1, x_2} P^M(x_1, x_2, x_1^{(i)}, x_2^{(i)}), \quad (23)$$

and the results of x_1 and x_2 in each iteration can be written separately as

$$[x_1^{(i+1)}]_n = \frac{[q_1(x_1^{(i)}, x_2^{(i)})]_n}{[p_1(x_1^{(i)}, x_2^{(i)})]_n}, \quad (24a)$$

$$[x_2^{(i+1)}]_n = \frac{[q_2(x_1^{(i)}, x_2^{(i)})]_n}{[p_2(x_1^{(i)}, x_2^{(i)})]_n}. \quad (24b)$$

Table 2 gives the specific steps of the proposed algorithm RTEA. Note that, in this algorithm, the computation of r_0 , r_1 and r_2 can be implemented directly with convolution, and the rest steps are element-wised independent, where practically multi-threading computation can be directly adopted. The systematic structure of the proposed RTEA for fault detection of rolling element bearings is presented in Figure 2. Since this algorithm is derived using MM procedure, the convergence can be guaranteed and converges to the optimal minimizer in this case. The detailed proof of MM for convex problems can be found in Ref. [28, Chapter 10].

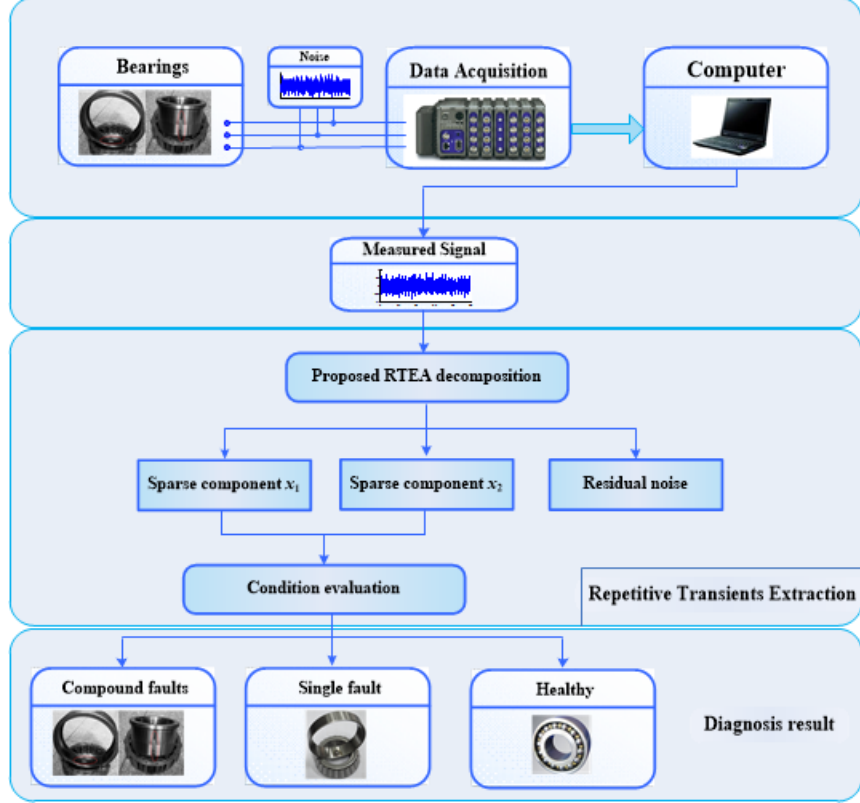


Figure 2: Procedure of the proposed RTEA for fault detection of bearings.

5 Synthetic data example

Figure 3 shows the synthetic test signal. It consists of two sparse transient sequences exhibiting distinct periods, where $T_1 = 32$ and $T_2 = 53$ samples. In this example, each transient is generated by adding a random number of sinusoids with random frequencies and initial phases. More specifically, each transient can be written as

$$g(n) = \sum_{j=1}^J A_j \sin(\omega_j n + \theta_j), \quad n \in \{0, 1, 2 \dots 9\}, \quad (25)$$

where $1 \leq J \leq 10$ is a random integer, and for each j , A_j is a random amplitude, ω_j is a random frequency, and θ_j is a random phase. The sequences of transients are shown in Figure 3(a) and Figure 3(b), and their summation is shown in Figure 3(c).

In this example, the parameters are set to $N_{11} = N_{21} = K_0 = 3$ and $M = 4$, to determine the binary arrays b_1 and b_2 by the priori known periods. The root-mean-square error (RMSE) is used as an evaluation metric. Moreover, the penalty functions are selected by Proposition 1 assuring the problem is strictly convex. Figure 4 shows the results from the proposed method, and Figure 5

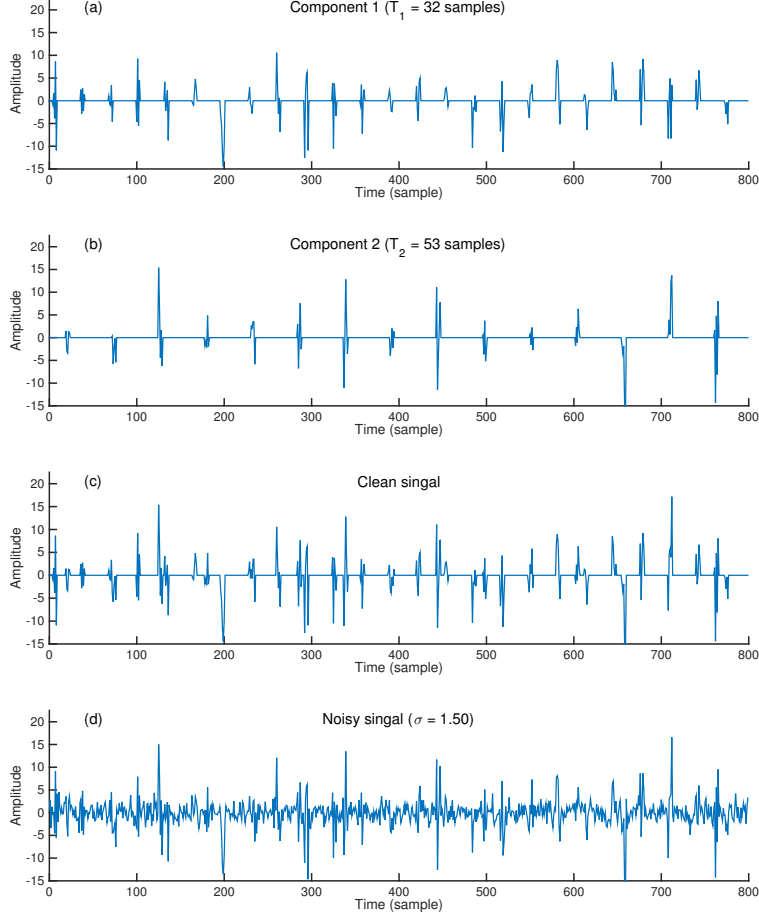


Figure 3: Example 1: Test signal. (a) Component 1 with period $T_1 = 32$ samples. (b) Component 2 with period $T_2 = 53$ samples. (c) Summation of the two components. (d) Noisy observation.

shows the convergence behavior of the proposed algorithm.

As a comparison, considering another problem, based strictly on MCA, whose objective function is

$$\arg \min_{x_1, x_2} \frac{1}{2} \|y - (x_1 + x_2)\|_2^2 + \sum_{i \in \{1, 2\}} \lambda_i \Phi(x_i; b_i). \quad (26)$$

Note that (26) has one regularization term less than (12), where the global sparsity is not promoted by function R .

Although non-convex penalties can help to promote sparsity, this will break the convexity of problem (26), so that a global optimal solution is not assured. Moreover, experiment results find that convex formulation (12) obtains as good a performance as the non-convex formulation (26). As a consequence, the formulation (12) is preferred, because there is no need to sacrifice convexity.

In addition, the proposed method with non-convex formulation is also evaluated, where a_1

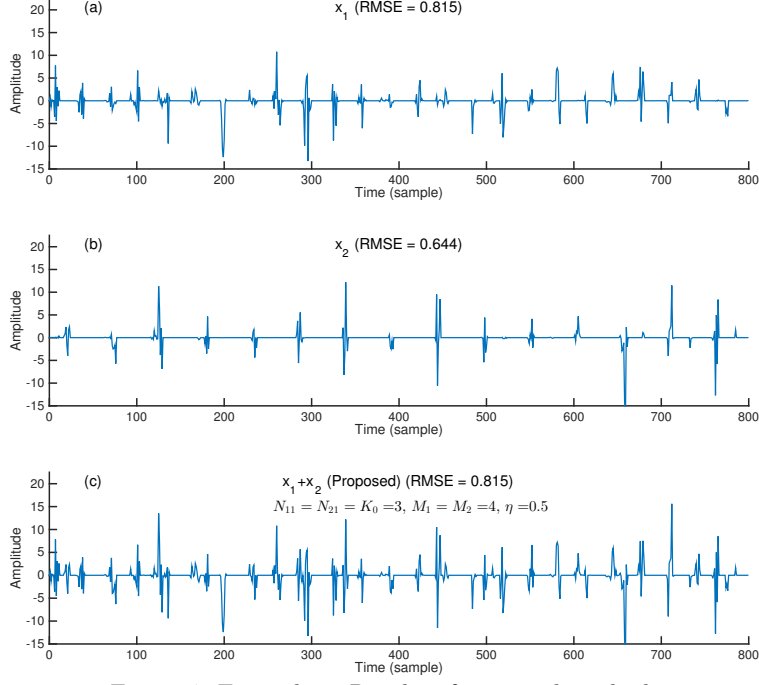


Figure 4: Example 1: Results of proposed method.

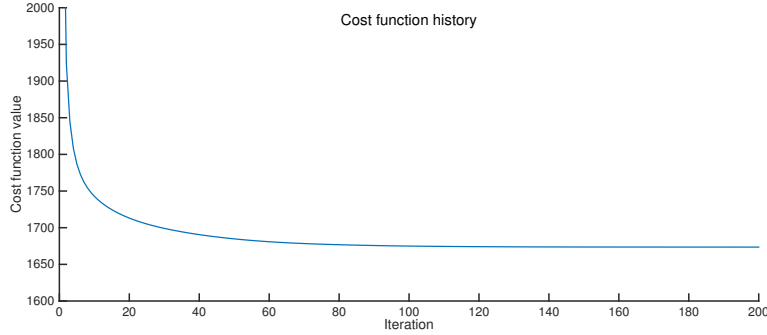


Figure 5: Example 1: Cost function history.

and/or a_2 are greater than zero. In this case, losing the convexity, the resulting sparsity can be even further promoted, then the formulation (12) is still preferred to the MCA formulation (26).

Further comparisons to some denoising methods are also presented. Firstly, wavelet-based denoising method is adopted to the test signal, more specifically, a 6-scale undecimated wavelet transform [15] using Haar wavelet filter. For denoising, hard-thresholding is applied and the threshold value is chosen by 3σ -rule for each subband. As shown in Figure 7(a), the denoising results adheres the shape of impulse response of the wavelet, where some of the signal does not have a zero baseline.

Secondly, the non-convex regularized fused lasso (FL) approach proposed in [6] is adopted, which is an improved version of conventional FL [41], allowing one regularizer to be non-convex

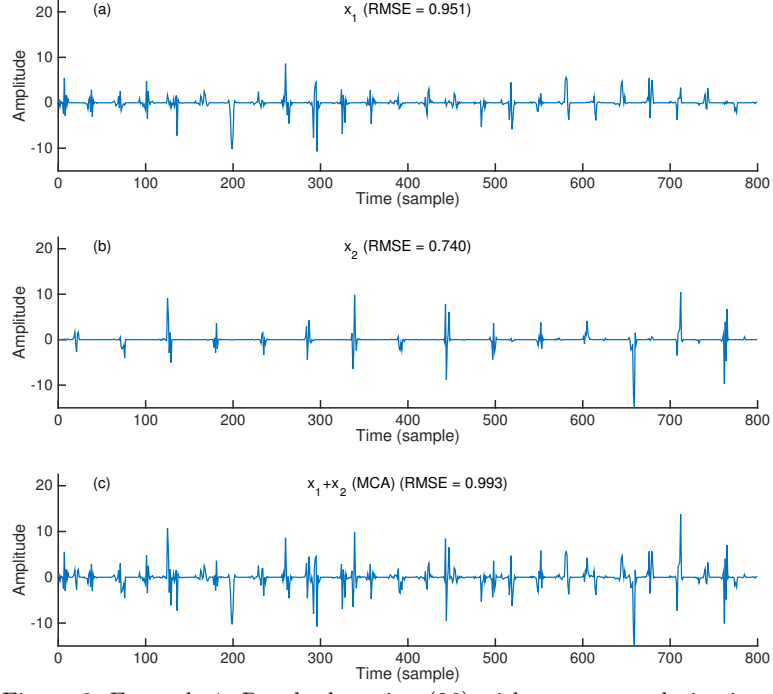


Figure 6: Example 1: Results by using (26) with convex regularizations.

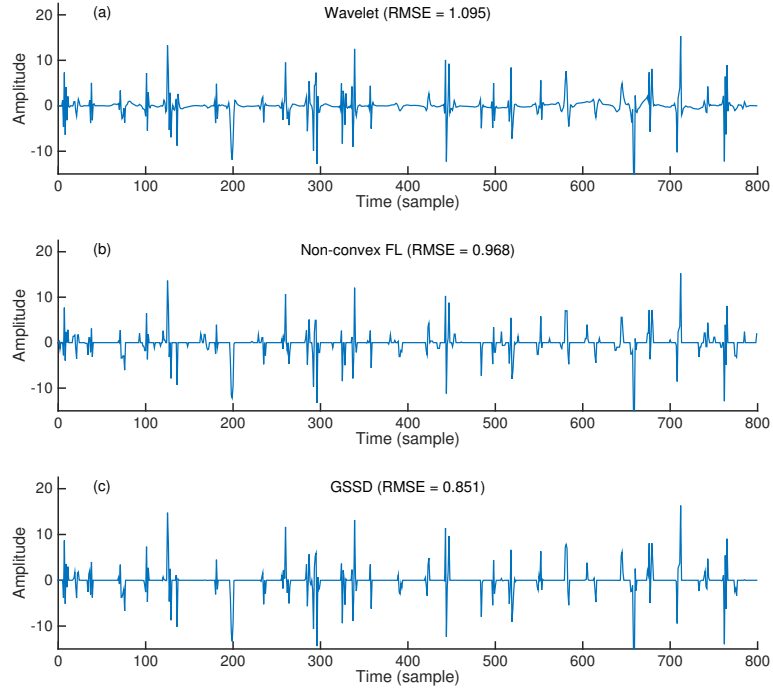


Figure 7: Example 1: Results by using some popular and/or state-of-the-art methods. (a) Undecimated wavelet based denoising. (b) Fused lasso with non-convex regularization. (c) group-sparse signal denoising (GSSD) with group size 3.

and preserving the global convexity. The result is shown in Figure 7(b). Thirdly, group-sparse signal denoising (GSSD) also known as OGS with non-convex regularization [13] is adopted. More specifically, ‘atan’ penalty with group size to be 3 samples is used, and the regularization parameter

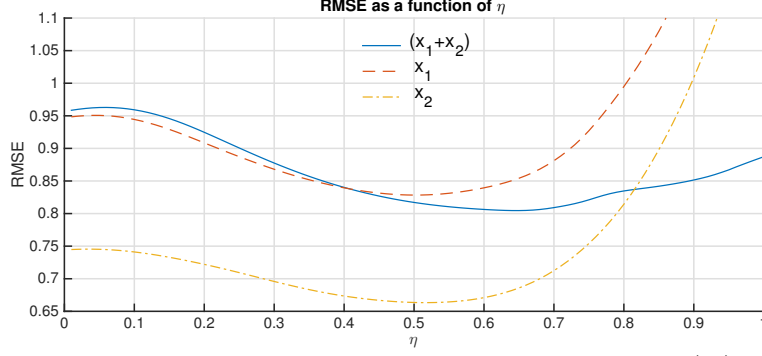


Figure 8: Example 1: RMSE values as a function of η using (27).

is chosen to optimize the RMSE. The result is shown in Figure 7(c). Both of the above methods also have worse recoveries in terms of RMSE. In addition, these methods are all denoising methods only, they cannot decompose the signal into two distinct sequences of transients, each exhibiting its own period.

5.1 Parameter selection

Table 3: Selection of β_i for $i = 0, 1, 2$.

$M_i \backslash N_{i1}$	1	2	3	4
1 (K_0)	3.700	1.700	1.150	0.925
2	1.700	0.850	0.625	0.475
3	1.150	0.625	0.450	0.375
4	0.925	0.475	0.375	0.325

In Section 4.1 of [24], the schemes to set the binary weight array b (14) and the regularization parameter λ for POGS problem have been discussed in detail. Moreover, a look-up table has been given as a guide to choose the regularization parameter, where using the given multiplier in [24, Table 3], the regularization parameter can be chosen by $\lambda = \beta\sigma_w$, where σ_w is the deviation of the additive noise. Here, the table is quoted with a slight change of notation.

As an extension of POGS with simultaneously decomposition, this parameter selection scheme can be used with a modification wherein the weights are shared among the three regularizers. In this case, the regularization parameters $\{\lambda_0, \lambda_1, \lambda_2\}$ are suggested to be set by

$$\lambda_0 = \eta\beta_0 \tag{27a}$$

$$\lambda_i = 0.5(1 - \eta)\beta_i \text{ for } i = 1, 2 \tag{27b}$$

where $0 < \eta < 1$ is a parameter to balance sparsity of the sum $(x_1 + x_2)$ and sparsity of the individual signal component x_i .

To run the method, firstly the binary weight arrays b_1, b_2 need to be set by (14) using the priori periods. Then β_i can be determined by Table 3 using N_{i1} and M_i for $i = 1, 2$. Note that it is necessary to chose $K_0 = \min\{N_{11}, N_{21}\}$ to induce group-sparsity. Hence in practice K_0 is not necessary to be chosen, and so as β_0 .

In practice, the parameters are set to be $K_0 = N_{11} = N_{21}$, and $M_1 = M_2 = 4$. The regularization parameters can be straight-forwardly determined by Table 3 and (27).

Setting parameter η . The parameter $0 < \eta < 1$ in (27) is to balance the sparsity, as mentioned above. As special cases:

1. If $\eta \rightarrow 1$, then $\lambda_1 = \lambda_2 \rightarrow 0$, and the approach promotes sparsity of $x_1 + x_2$, but leads to $x_1 = x_2$.
2. If $\eta \rightarrow 0$, then the problem (12) reduces to (26), which is a conventional MCA problem. It is able to sepearate the components, but promotes sparsity weakly.

To obtain both of the benefits, through experiments, parameter η is suggested to be set around 0.5.

Figure 8 illustrates the RMSE values of x_1, x_2 and $(x_1 + x_2)$ as functions of η . As shown, when η is greater than 0.5, although the RMSE of the global signal $(x_1 + x_2)$ is low, the performance of the decomposition is worse. Moreover, η should not be too small, for then the results will be similar to conventional MCA shown in Figure 6.

6 Engineering Examples

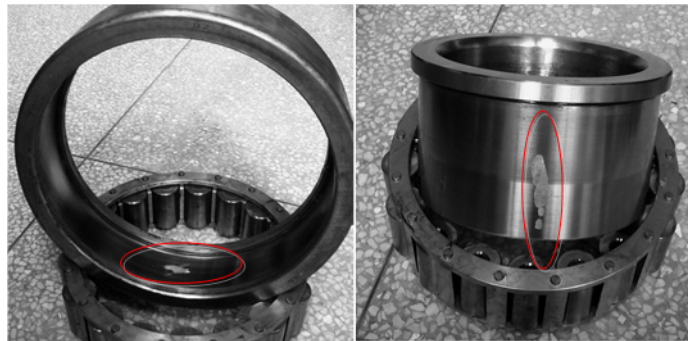


Figure 9: Outer race defect (left) and inner race (right) in the bearing.

Table 4: Parameters of 552732QT bearing

Inner Race (mm)	Outer Race(mm)	Roller (mm)	Number of rollers	Contact angle (degree)
160	290	34	17	0

In this section, the proposed RTEA is applied to analyze vibration signals collected from a rolling element bearing with compound defects on a locomotive, using a SONY EX data acquisition system operating at a sampling rate of $f_s = 12.8$ kHz. The locomotive bearing with faults in inner and out races is shown in Figure 9. The bearing (552732QT) parameters are given in Table 4.

6.1 Compound faults detection

In this example, the vibration signals were collected at a constant shaft speed of 360 r/min. Thus, based on the geometric parameters and rotational frequency, the ball-pass frequency of the outer race is about 43.3 Hz, and that of the inner race is about 58.7 Hz.

In practice, the regularization parameters in (12) can be estimated from the noise level, or the deviation of the vibration signal without any fault under an approximately identical experiment environment. Moreover, when the healthy data is not available, the ‘noise’ level can still be determined by the formula

$$\hat{\sigma} = \text{MAD}(y)/0.6745 \quad (28)$$

which is a conventional estimator of noise level used for wavelet-based denoising [20], where MAD is the median absolute deviation defined as

$$\text{MAD}(y) := \text{median}(|y_n - \text{median}(y)|). \quad (29)$$

In this example, the formula (28) is used to estimate σ directly from the observation data, and then that value is used to choose the regularization parameters from Table 3.

Figure 10 shows the results using RTEA, where x_1 is the transient sequence generated by the fault in the outer race, and x_2 is the transient sequence generated by the fault in the inner race. During the recording, the outer race is stable and the inner race is rotating. Hence, the amplitude of transients in x_2 (inner race fault) exhibit modulating effect.

To further reveal the characteristic frequencies, the Hilbert envelope spectrums of the extracted components are shown in Figure 11. The smoothed (by lowpass filtering) profiles of the Hilbert

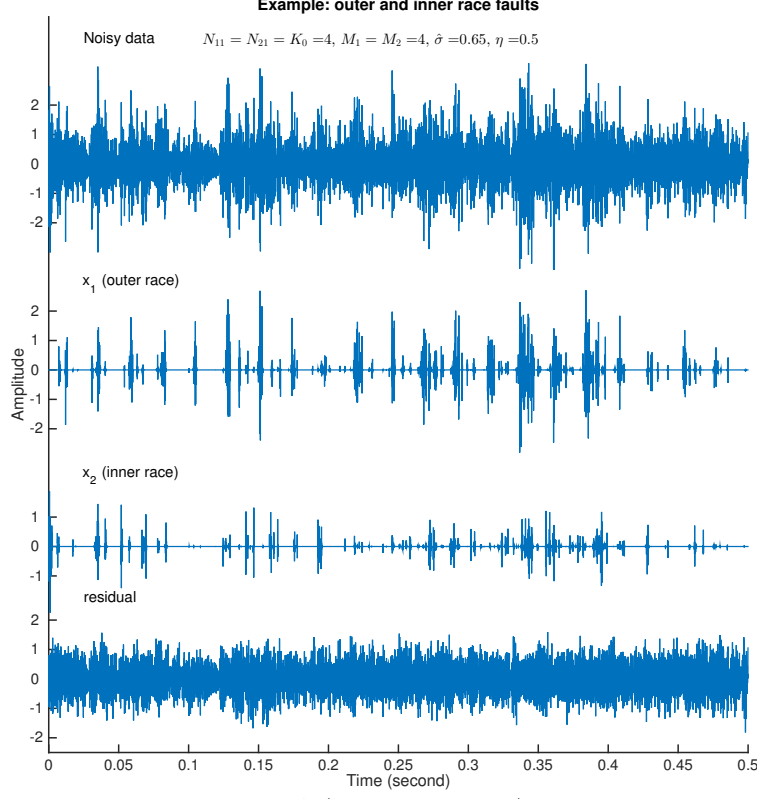


Figure 10: Example 2: Output of RTEA (proposed method) from data with compound faults.

envelope spectrum is also presented to indicate the characteristic frequencies more clearly. Figure 11 shows that the characteristic frequencies of outer and inner race are at about 45 and 60 Hz, and their harmonics are revealed by the peaks of the smoothed profiles. Thus, the fault features of the two defects are clearly extracted by the proposed approach.

As a comparison, we include the Welch's estimate of the cyclic spectral coherence for the analysis of cyclostationary signals [2, 3, 4]. The result is illustrated in Figure 12. Note that, the signal length is 6400, the window length is 32, the signal is divided into 14 overlapping blocks and the block overlap is 21 with a Hanning window, which is suggested in Ref. [2]. When observing the coherence, the cyclic frequency of the outer race (approximate 43.3 Hz) and its harmonic components can be identified. However, the characteristic frequencies of inner race 58.7 Hz and its harmonic components cannot be observed from Figure 12. This may be caused by the fact that the component of inner race in this case is weak.

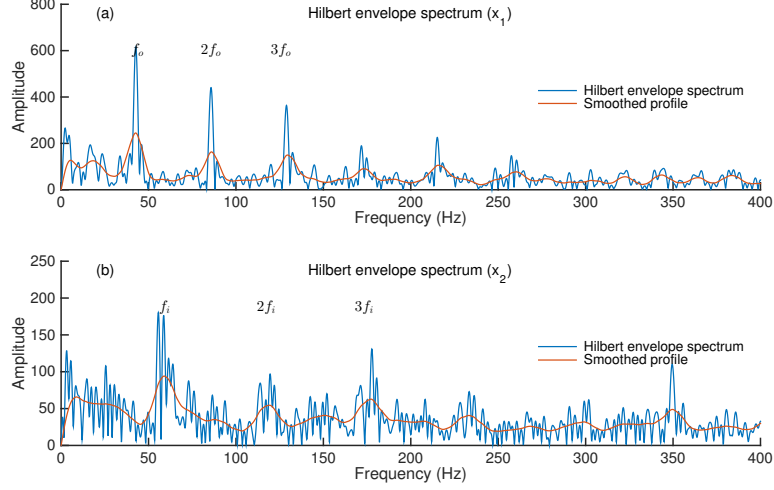


Figure 11: Example 2: Hilbert envelope spectra of the extracted repetitive transient sequences.

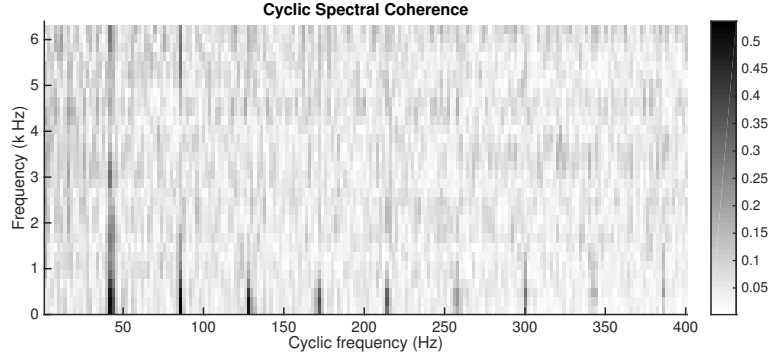


Figure 12: Example 2: Cyclic spectral coherence of the test signal given in Figure 10.

6.2 Example 3: Single fault detection

In this example, the proposed method also works when the bearing has only one fault. Figure 13 shows the measured data and the results from the bearing in Figure 9, but there was only one defect on the outer race. The acceleration signals were collected at a constant shaft speed of 481 r/min. Thus, the characteristic fault frequencies of the outer and inner races are about 57.8 Hz and 78.4 Hz respectively.

Using the given information of the periods, the repetitive transient sequences x_1 , x_2 and the residual noisy signals are extracted simultaneously from the measured vibration data, where x_1 is corresponds to outer race defect. Repetitive transients can be observed from the extracted x_1 , where the fault frequency can be directly observed around 58 Hz, because there are about 29 sparse transients evenly distributed with the 0.5 second. The extracted characteristic frequency 58 Hz is approximately in accordance with the outer race fault frequency of 57.8 Hz.

Moreover, component x_2 in Figure 13 is almost purely zero and it exhibits almost no repetitive

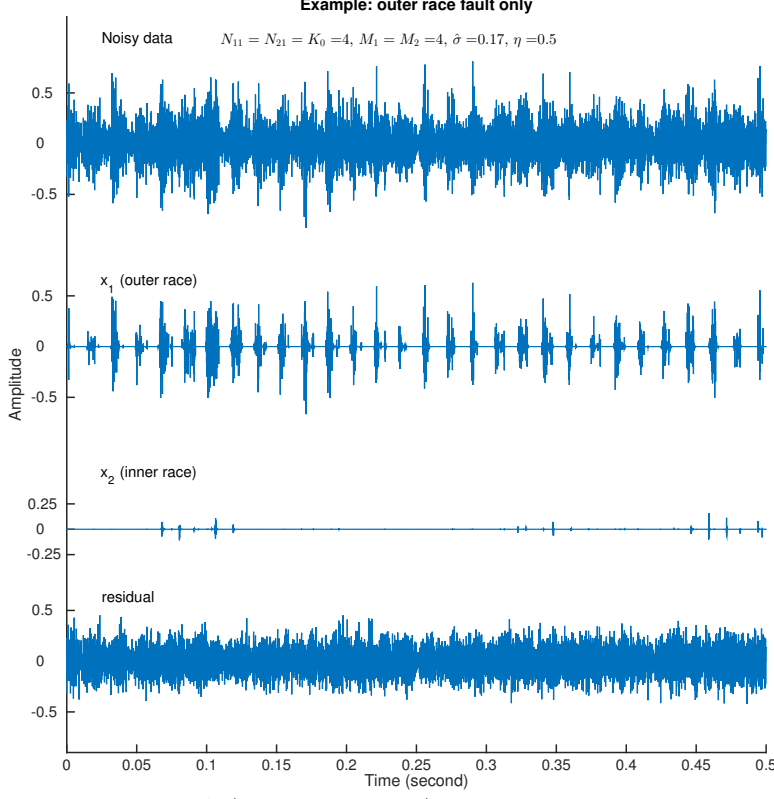


Figure 13: Example 3: Output of RTEA (proposed method) from data with outer race fault only. Note that the Y-axis has a different scale to the previous example.

transients. This implies that the status of the inner race is healthy.

7 Conclusion

This paper proposes a novel approach for the extraction of repetitive transients with group-sparse structure in vibration signals, and for detecting faults in rolling element bearings. To simultaneously extract both sparse components and perform denoising, a repetitive group-sparsity based optimization problem is formulated. To solve the problem, a computationally efficient iterative algorithm, termed repetitive transients extraction algorithm(RTEA) is derived. The non-convex penalty function is used to strongly promote sparsity, and a condition is given so that the objective function is strictly convex. Moreover, for practical problems, an approach to select regularization and non-convexity parameters is provided.

Appendices

A Proof of Proposition 1

Proof. The first two terms of the objective function in (12) can be rewritten as

$$F(u) = \frac{1}{2}\|y - u\|_2^2 + \lambda_0 \sum_n \phi_\epsilon \left(\left[\sum_{k=0}^{K_0-1} [u]_{n+k}^2 \right]^{1/2}; a_0 \right) \quad (\text{A.1})$$

where $u \in \mathbb{R}^N$ and $u - (x_1 + x_2) = \mathbf{0}$. Note that $F(u)$ is exactly identical to GSSD problem {see Equation (20) in [13, Theorem 1]}. Adopting Theorem 1 and Corollary 2 in [13], it can be shown that: when

$$0 \leq a_0 < \frac{1}{K_0 \lambda_0}, \quad (\text{A.2})$$

$F(u)$ is strictly convex. Moreover, it is immediate that when $a = 0$, function (11) is convex. As a consequence, an equivalent problem of (12) is considered, namely

$$\begin{aligned} \{u^{\text{opt}}, x_1^{\text{opt}}, x_2^{\text{opt}}\} &= \arg \min_{u, x_1, x_2} F(u) + \sum_{i \in \{1, 2\}} \lambda_i \Phi(x_i; b_i) \\ &\text{such that } u - x_1 - x_2 = \mathbf{0}, \end{aligned} \quad (\text{A.3})$$

which satisfies the convexity condition of equality constrained problem in [11], and which implies that: when (12) satisfies (15), it is a convex problem. \square

B Derivation of majorizing function R (13)

Since ϕ_ϵ^{M} in (7) majorizes ϕ_ϵ , an upper bound of function R can be found as

$$\begin{aligned} \bar{R}(x_1, x_2, v; a_0) &= \sum_n \phi_\epsilon^{\text{M}} \left(\left[\sum_k [x_1 + x_2]_{n+k}^2 \right]^{1/2}, v_n; a_0 \right) \\ &= \sum_n \left\{ \frac{1}{2\psi(v_n; a_0)} \sum_k [x_1 + x_2]_{n+k}^2 \right\} + C \end{aligned} \quad (\text{B.1})$$

where C is a constant does not depend on x_1 or x_2 , and for any $v \in \mathbb{R}^N$,

$$R(x_1, x_2; a_0) \leq \bar{R}(x_1, x_2, v; a_0). \quad (\text{B.2})$$

After algebraic manipulations, \bar{R} can be further expressed as

$$\bar{R}(x_1, x_2, v(z); a_0) = \frac{1}{2} \sum_n r_0(n, z) [x_1 + x_2]_n^2 + C, \quad (\text{B.3})$$

where vector v is dependent on another vector $z \in \mathbb{R}^N$, as

$$[v(z)]_n = \left[\sum_{k=0}^{K_0-1} [z]_{n+k}^2 \right]^{1/2}, \quad (\text{B.4})$$

and $r_0 : \mathbb{Z} \times \mathbb{R}^N \rightarrow \mathbb{R}$, is defined by

$$r_0(n, z) := \sum_{j=0}^{K_0-1} \frac{1}{\psi([v(z)]_{n-j}; a_0)} = \sum_{j=0}^{K_0-1} \frac{1}{\psi\left(\left[\sum_k [z]_{n-j+k}^2\right]^{1/2}; a_0\right)} \quad (\text{B.5})$$

which is similar to Equation (37) of [13].

Note that, in this case the inequality of (B.2) is still valid and the equality holds when

$$z = x_1 + x_2. \quad (\text{B.6})$$

Moreover, considering a simple inequality that: for any $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$,

$$(\alpha_1 + \alpha_2)^2 \leq (\alpha_1 + \alpha_2)^2 + [(\alpha_1 - \beta_1) - (\alpha_2 - \beta_2)]^2, \quad (\text{B.7})$$

whose right side can be re-written as

$$2\alpha_1^2 + 2\alpha_2^2 - 2(\beta_1 - \beta_2)\alpha_1 - 2(\beta_2 - \beta_1)\alpha_2 + (\beta_1 - \beta_2)^2, \quad (\text{B.8})$$

where α_1 and α_2 are de-coupled and the equality holds when $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$. Then, using the inequality (B.7) element-wise, an upper bound of \bar{R} can be found as

$$\begin{aligned} & R^M(x_1, x_2, z_1, z_2; a_0) \\ &= \sum_n \left\{ r_0(n, z_1 + z_2) ([x_1]_n^2 + [x_2]_n^2 - [z_1 - z_2]_n [x_1]_n - [z_2 - z_1]_n [x_2]_n) \right\} + C(z_1, z_2), \end{aligned} \quad (\text{B.9})$$

where $C(z_1, z_2)$ is a constant only dependent on z_1 and z_2 . Note that since (B.9) is an upper bound

of \bar{R} , consequently

$$R^M(x_1, x_2, z_1, z_2; a_0) \geq R(x_1, x_2; a_0). \quad (\text{B.10})$$

Furthermore, when $z_1 = x_1$ and $z_2 = x_2$, using (B.7) it can be seen that $R^M(x_1, x_2, z_1, z_2; a_0) = \bar{R}(x_1, x_2, v(z_1 + z_2); a_0)$, and this implies that

$$R^M(x_1, x_2, x_1, x_2; a_0) = R(x_1, x_2; a_0). \quad (\text{B.11})$$

Therefore, equation (B.9) is a majorizer of function R in (13).

References

- [1] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mech. Syst. Signal Process.*, 21(7):2933–2945, 2007.
- [2] J. Antoni. Cyclic spectral analysis in practice. *Mech. Syst. Signal Process.*, 21(2):597–630, 2007.
- [3] J. Antoni. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. 304(35):497–529, 2007.
- [4] J. Antoni. Cyclostationarity by examples. *Mech. Syst. Signal Process.*, 23(4):987–1036, 2009.
- [5] J. Antoni and R. B. Randall. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mech. Syst. Signal Process.*, 20(2):308–331, 2006.
- [6] I. Bayram, P.-Y. Chen, and I. W. Selesnick. Fused lasso with a non-convex sparsity inducing penalty. In *Proc. ICASSP 2014*, pages 4156–4160, May 2014.
- [7] G. Bin, J. Gao, X. Li, and B. S. Dhillon. Early fault diagnosis of rotating machinery based on wavelet packets -empirical mode decomposition feature extraction and neural network. *Mech. Syst. Signal Process.*, 27:696–711, 2012.
- [8] T. Boukra, A. Lebaroud, and G. Clerc. Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation. *IEEE Trans. Ind. Electron.*, 60(9):4034–4042, 2013.

- [9] R. Boustany and J. Antoni. A subspace method for the blind extraction of a cyclostationary source: Application to rolling element bearing diagnostics. *Mech. Syst. Signal Process.*, 19(6):1245–1259, 2005.
- [10] R. Boustany and J. Antoni. Blind extraction of a cyclostationary signal using reduced-rank cyclic regression: a unifying approach. *Mech. Syst. Signal Process.*, 22(3):520–541, 2008.
- [11] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [12] B. Chen, Z. Zhang, C. Sun, B. Li, Y. Zi, and Z. He. Fault feature extraction of gearbox by using overcomplete rational dilation discrete wavelet transform on signals measured from vibration sensors. *Mech. Syst. Signal Process.*, 33:275–298, 2012.
- [13] P.-Y. Chen and I. W. Selesnick. Group-sparse signal denoising: Non-convex regularization, convex optimization. *IEEE Trans. Signal Process.*, 62(13):3464–3478, July 2014.
- [14] S. Chen and D. L. Donoho. Basis pursuit. In *1994 Conference Record of the Twenty-Eighth Asilomar, Conference on Signals, Systems and Computers, 1994.*, volume 1, pages 41–44, October 1994.
- [15] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In *Wavelet and statistics*, pages 125–150. Springer-Verlag, 1995.
- [16] L. Cui, J. Wang, and S. Lee. Matching pursuit of an adaptive impulse dictionary for bearing fault diagnosis. *J. Sound Vib.*, 333(10):2840–2862, 2014.
- [17] L. Cui, N. Wu, C. Ma, and H. Wang. Quantitative fault analysis of roller bearings based on a novel matching pursuit method with a new step-impulse dictionary. *Mech. Syst. Signal Process.*, 68:34–43, 2016.
- [18] L. Cui, Y. Zhang, F. Zhang, J. Zhang, and S. Lee. Vibration response mechanism of faulty outer race rolling element bearings for quantitative analysis. *J. Sound Vib.*, 364:67–76, 2016.
- [19] Y. Ding and I. W. Selesnick. Sparsity-based correction of exponential artifacts. *Signal Process.*, 120:236–248, March 2016.
- [20] D. Donoho, I. Johnstone, and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1993.

- [21] Z. Feng, M. Liang, and F. Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. *Mech. Syst. Signal Process.*, 38(1):165–205, 2013.
- [22] M. Figueiredo, J. Bioucas-Dias, and R. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Trans. Image Process.*, 16(12):2980–2991, December 2007.
- [23] Q. He and X. Wang. Time–frequency manifold correlation matching for periodic fault identification in rotating machines. *J. Sound Vib.*, 332(10):2611–2626, 2013.
- [24] W. He, Y. Ding, Y. Zi, and I. W. Selesnick. Sparsity-based algorithm for detecting faults in rotating machines. *Mech. Syst. Signal Process.*, 72-73:46–64, May 2016.
- [25] W. He, Y. Zi, B. Chen, S. Wang, and Z. He. Tunable Q-factor wavelet transform denoising with neighboring coefficients and its application to rotating machinery fault diagnosis. *Sci. China Technol. Sci.*, 56(8):1956–1965, 2013.
- [26] W. He, Y. Zi, B. Chen, F. Wu, and Z. He. Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform. *Mech. Syst. Signal Process.*, 54-55:457–480, 2015.
- [27] H. Jiang, J. Chen, G. Dong, T. Liu, and G. Chen. Study on Hankel matrix-based SVD and its application in rolling element bearing fault diagnosis. *Mech. Syst. Signal Process.*, 52:338–359, 2015.
- [28] K. Lange, D. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *J. of Comp. Graph. Statist.*, 9:1–20, 2000.
- [29] Y. Lei, Z. He, Y. Zi, and X. Chen. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique. *Mech. Syst. Signal Process.*, 22(2):419–435, 2008.
- [30] Y. Lei, J. Lin, Z. He, and M. J Zuo. A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.*, 35(1):108–126, 2013.
- [31] Z. Li, X. Yan, C. Yuan, Z. Peng, and L. Li. Virtual prototype and experimental research on gear multi-fault diagnosis using wavelet-autoregressive model and principal component analysis method. *Mech. Syst. Signal Process.*, 25(7):2589–2607, 2011.

- [32] M. Liang and I. Soltani Bozchalooi. An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection. *Mech. Syst. Signal Process.*, 24(5):1473 – 1494, 2010.
- [33] A. Malhi and R. X. Gao. PCA-based feature selection scheme for machine defect classification. *IEEE Trans. Instrum. Meas.*, 53(6):1517–1525, 2004.
- [34] Y. Qin, Y. Mao, and B. Tang. Vibration signal component separation by iteratively using basis pursuit and its application in mechanical fault detection. *J. Sound Vib.*, 332(20):5217–5235, 2013.
- [35] R. B. Randall and J. Antoni. Rolling element bearing diagnostics – a tutorial. *Mech. Syst. Signal Process.*, 25(2):485–520, 2011.
- [36] I. W. Selesnick and I. Bayram. Sparse signal estimation by maximally sparse convex optimization. *IEEE Trans. Signal Process.*, 62(5):1078–1092, 2014.
- [37] I. W. Selesnick, A. Parekh, and I. Bayram. Convex 1-d total variation denoising with non-convex regularization. *IEEE Signal Processing Letters*, 22(2):141–144, 2015.
- [38] W. A. Smith and R. B. Randall. Rolling element bearing diagnostics using the case western reserve university data: A benchmark study. *Mech. Syst. Signal Process.*, 64:100–131, 2015.
- [39] J.-L. Starck, M. Elad, and D. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132:287–348, 2004.
- [40] W. Su, F. Wang, H. Zhu, Z. Zhang, and Z. Guo. Rolling element bearing faults diagnosis based on optimal morlet wavelet filter and autocorrelation enhancement. *Mech. Syst. Signal Process.*, 24(5):1458–1472, 2010.
- [41] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society Series B*, pages 91–108, 2005.
- [42] Z. Wang, J. Chen, G. Dong, and Y. Zhou. Constrained independent component analysis and its application to machine fault diagnosis. *Mech. Syst. Signal Process.*, 25(7):2501–2512, 2011.
- [43] R. Yan and R. X. Gao. Harmonic wavelet-based data filtering for enhanced machine defect identification. *J. Sound Vib.*, 329(15):3203–3217, 2010.

- [44] R. Yan, R. X. Gao, and X. Chen. Wavelets for fault diagnosis of rotary machines: A review with applications. *Signal Process.*, 96:1–15, 2014.
- [45] H. Yang, J. Mathew, and L. Ma. Fault diagnosis of rolling element bearings using basis pursuit. *Mech. Syst. Signal Process.*, 19(2):341–356, 2005.
- [46] X. Zhang and J. Zhou. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. *Mech. Syst. Signal Process.*, 41(1):127–140, 2013.